The Social Geography of American Medicine

Abstract: Health care exhibits wide geographic variation. Because care is driven by the behavior of physicians, regional differences in its cost and quality are thought to reflect local cultures of medical practice, but the mechanisms producing and maintaining these putative cultures remain unclear. Local billing and prescribing cultures may be reflected in physician referral networks, but the complexity of U.S. health care has made it difficult to study such networks on a nationwide basis. Here we combine comprehensive, longitudinal, publicly available data on physician billing, prescribing, and patient-sharing behavior to characterize the "social geography" underlying area variations in health care. Focusing on six measures of billing and prescribing intensity, we investigate the clustering of physicians' behavior on the basis of their social proximity in patient-sharing networks, finding that both social influence and selection contribute to the clustering observed. Our results have implications for efforts to improve the value and quality of health care, and highlight the potential of publicly available administrative data in promoting transparency in research and public affairs.

Table of Contents

Glossary

CMS: Centers for Medicare & Medicaid Services
GDP: Gross domestic product
HRR: Hospital referral region
NPI: National Provider Identifier

Introduction

The increasing availability of detailed administrative records has placed vast amounts of previously inaccessible data in the public domain. The proliferation of such "big data" sources presents opportunities to answer questions of broad societal interest (1-3), with an open approach ideal for the investigation of sensitive topics, such as the potential misuse of public funds.

As new medical technologies proliferate, health care spending increases unsustainably worldwide. The United States in particular spent 17.4% of its GDP on health in 2013 (4), far more than any other country, but with outcomes no better than those of its peers (5, 6).

Even within the U.S., however, the intensity of medical practice is far from homogeneous. As early as 1973, Wennberg & Gittelsohn characterized dramatic regional variations in medical utilization and spending (7). High-spending areas incur costs 50% greater than low-spending ones (8), and have been associated with inferior quality and outcomes (9, 10). Though patient needs and demographics differ by region, the primary driver of practice variation is physician behavior (11). Some physicians consistently deliver high-quality care at 80% the cost of their peers (12). Conversely, as much as 17% of U.S. health care spending is driven by the behavior of high-intensity providers making decisions unsupported by clinical evidence (11).

What drives the behavior of individual physicians, and accounts for the dramatic geographic variations therein? Though physicians differ in the extent to which they adhere to evidence-based clinical guidelines (11), much of the variation in medical spending and quality is driven by clinical "grey areas" (13, 14), for which no unambiguous professional guidelines exist. Physicians have wide discretion in applying costly diagnostic and therapeutic technologies, as well as in how they bill for their work. Since billing is not taught in medical school or residency, and is not subject to evidence-based guidelines, billing behavior may be particularly susceptible to social and cultural influence. Young physicians learn primarily by watching others (6), and even established physicians who move to regions with more or less aggressive billing practices modify their own behavior to align with the local norm (15).

Social-professional relationships among physicians are thus a prime candidate conduit for the diffusion of clinical and billing practices, but the enormous structural complexity of U.S. health care has made it difficult to study such networks on a nationwide basis. To our knowledge, the most complete studies of physician networks to date (16) have analyzed at most 51 of the 306 (17%) U.S. "hospital referral regions" (HRRs).

In this study we combine recently released, publicly available data on physician billing, prescribing, and patient-sharing behavior. These data are nationally complete, containing information on billions of medical services (2.5 billion each in 2012 and 2013) and prescriptions (1.2 billion in 2012, 1.4 billion in 2013), and hundreds of millions of patient-sharing relationships (55 million each in 2012 and 2013), for providers billing Medicare (America's national health insurance program for citizens aged 65 and older, or with certain disabilities). We use these data to construct comprehensive longitudinal networks of Medicare providers. We augment these networks with complete outpatient billing and prescribing data for providers billing Medicare in 2012 and 2013. For all providers billing for at least 100 office visits in 2013, we construct a panel of specialty-adjusted billing intensity measures. We then calculate, for all 112 million pairs of eligible providers sharing the same HRR, metrics of specialty-adjusted similarity in practice intensity, and compare dyads' similarity to their members' proximity in the network and the volume of patients they share, in order to characterize the social patterning of physician behavior throughout the United States. In longitudinal analyses, we identify potential mechanisms productive of clustering in physician behavior, with implications for efforts to improve the value and quality of care.

Methods

Description of Data Sets: Patient-Sharing Data. Beginning in 2012, in response to a Freedom of Information Act request, the Centers for Medicare & Medicaid Services (CMS) has publicly released data on hundreds of millions of patient-sharing relationships between Medicare providers, representing many billions of instances of shared patients, for the years 2009, 2010, 2011, 2012, 2013, 2014 and 2015.

Each patient-sharing relationship is characterized by the National Provider Identifier (NPI) numbers of two providers, a count of shared patient transactions over a 30-day window, and the total number of unique patients shared over that window. A "shared patient transaction" is counted when Provider *i* sees (and bills for) a given patient first, and Provider *j* sees (and bills for) that same patient later, within 30 days of Provider *i* seeing the patient. The data also identify the number of unique patients involved in a patient-sharing relationship: a relationship between two providers with a "shared patient transaction count" of 20 could involve two providers each seeing 20 different patients once each within the same 30 day window, or each seeing the same patient 20 times within that same window.

To protect patient privacy, the data include only relationships involving at least 11 shared patients. Thus many "weak" patient-sharing ties are not observed in our data. In our longitudinal analyses, we therefore observe "tie formation" when a dyad sharing 0-10 patients in 2012 shares 11 or more in 2013, and "tie dissolution" when the reverse applies.

The 2013 patient-sharing data include 55,141,669 directed edges between pairs of providers sharing 11 or more patients within a 30-day window. The 2012 data 54,966,715 observations.

Description of Data Sets: Service and Billing Data. In June 2015, CMS released service and payment data for 956,293 providers billing Medicare Part B in 2013, as well as updated data for 925,328 providers billing Medicare Part B in 2012 (data on 880,576 of these providers was first released in April 2014).

These data contain information on the specific services for which providers billed Medicare Part B (2,194,068,109 in 2012; 2,189,252,757 in 2013) representing $199 billion in expenditures by Medicare and other payers for these two years. A full description of the data is available elsewhere (17).

At the level of provider (925,328 in 2012; 956,293 in 2013), these data contain a provider's name, practice location, gender, and specialty; the number of types of procedures for which the provider billed Medicare; the number of Medicare beneficiaries receiving the provider's services; and the total amounts billed by and paid to the provider that year.

At the level of provider-service (9,153,272 in 2012; 9,287,876 in 2013), the data contain the code and description of the service, the number of times it was performed by the provider, the number of unique beneficiaries receiving the service from that provider, and the average amounts billed and received by the provider for that service.

Description of Data Sets: Part D Prescribing Data. In April 2015, CMS publically released, for the first time, comprehensive data on all prescriptions billed to Medicare Part D in 2013 (18). For 1,049,381 providers writing prescriptions billed at least in part to Medicare, the data contain, among other information:

- The specific drugs prescribed by each provider
- For each drug a provider prescribes:
    - The number of patients for whom the drug was prescribed
    - The total number of prescription days (e.g., pills)
    - The costs associated with these prescriptions
    - The breakdown of beneficiaries by age, insurance status, and other variables
    - How often a provider prescribes brand-name vs. generic drugs

Analogous prescribing data for 2012 were obtained through ProPublica (19).

Description of Data Sets: Dartmouth Health Atlas. For HRR geographic boundary files, we use publicly available data from the Dartmouth Health Atlas (20).

Network construction. We use the patient-sharing data to construct comprehensive networks of Medicare providers in 2012 and 2013, in which pairs of physicians who shared at least 11 patients in a rolling 30-day window are connected by an edge.

To construct the provider networks for each period (2012, 2013), we selected only those patient-sharing relationships in which both providers were physicians practicing a specialty to or from which a referral is likely to signify a professional relationship between the "sending" and "receiving" physician. This method has been previously validated as a means of identifying real-world, physician-identified social relationships (*24*).

We admitted the following provider types:

- Internal Medicine
- Family Practice
- Cardiology
- Gastroenterology
- Pulmonary Disease
- Nephrology
- Hematology/Oncology
- General Practice
- Endocrinology
- Infectious Disease
- Medical Oncology
- Geriatric Medicine
- Hematology
- Osteopathic Manipulative Medicine
- Preventive Medicine
- Neurology

- Pediatric Medicine
- Rheumatology
- Allergy/Immunology
- Sports Medicine
- Physical Medicine and Rehabilitation
- Psychiatry
- Cardiac electrophysiology

We include all such providers billing for at least 100 established patient office visits in 2013.

We reduce the directed edges (in which the count of shared patient transactions signifies the number of times Provider i saw a patient within a 30 day window prior to Provider j seeing the same patient) to undirected edges (in which the count of shared transactions signifies the number of times either provider saw a patient within a 30 day window prior to the other provider seeing the same patient).

For all pairs of directly connected providers, each of whom billed Medicare for at least 100 office visits in 2013, we calculate four measures of edge strength:

1) The number of unique patients shared in 2013.

2) The number of patient-sharing instances in 2013 (i.e., the number of times both providers billed for a shared patient within a rolling 30 day window, equal to (1) if both providers billed for each shared patient on only one occasion).

3) A weighted measure of (1), where

$$w.dir_{ij} = \frac{p_{ij}}{\sum_k p_{ik}}$$

defines the weight of the directed edge from physician i to physician j, with p_{ij} denoting the number of patients referred from i to j, and k indexing all physicians to whom i refers patients,

$$w.dir_{ji} = \frac{p_{ji}}{\sum_k p_{jk}}$$

is defined analogously for the directed edge from $j \rightarrow i$, and

$$w.undir_{ij} = \frac{w.dir_{ij} + w.dir_{ji}}{2}$$

gives the average of the two directed measures, used as the undirected measure of edge strength. Thus a value of 1 signifies that providers i and j share patients exclusively with one another.

4) A weighted measure of (2), constructed analogously to (3).

The latter two measures serve to adjust for the volume of patients seen by the providers party to each dyad, so as to produce measures of edge strength more meaningfully comparable across providers sharing different overall volumes of patients (21-23).

Measures of billing intensity. We then use the billing and prescribing data to calculate a panel of six specialty-adjusted intensity measures for every provider (where "specialty" represents a physician's primary field of practice, e.g., cardiology, psychiatry, general practice), separately for 2012 and 2013.

We restrict our analysis to the 206,873 physicians of eligible specialty types billing for at least 100 routine office visits in 2013. Prescribing data is available for over 99% of these physicians.

Office visits for established patients are billed to Medicare at five levels of clinical intensity and concomitant reimbursement: the least complex and time-consuming visits are meant to be billed at level 1, which is the least well reimbursed, while the most complex and time-consuming visits are intended to be billed at level 5, and are reimbursed at higher rates.

For each physician, we calculate six measures of billing intensity, each normalized by the physician's specialty (such that a physician's score on any measure represents her *z-score* for that measure, compared to all other physicians practicing the same specialty), separately for 2012 and 2013:

1) Mean office visit intensity
2) Proportion of office visits billed at levels 4 or 5
3) Payment per beneficiary
4) Number of services billed per beneficiary
5) Proportion of prescriptions that are brand-name (vs. generic)
6) Prescription cost per beneficiary

We also aggregate physician intensity scores by hospital referral region (HRR), to study regional variation in our measures of clinical and billing intensity.

Measures (1) and (2) refer to "established patient office visits," which can be billed at five levels of "intensity" and concomitant reimbursement, from 1 (for a quick, routine visit) to 5 (for a complex and lengthy one), subject to guidelines but substantially at the billing physician's discretion.

Clustering of physician intensity: non-parametric approach. We then compute, for each of the 112,547,233 within-HRR dyads in the patient-sharing network (pairs of physicians, connected or not, sharing the same HRR in 2013), the similarity of the two physicians on each of the six intensity measures, and the geodesic distance (shortest path in the patient-sharing network) between them, in order to characterize the social clustering of billing and prescribing behavior.

For all analyses, we consider only within-HRR paths. Because HRRs were developed to study regional variation in medical practice and spending (24), using only within-HRR paths to investigate clustering of billing behaviors furnishes a conservative test: any clustering observed is net of aggregate geographic variation.

Using the unweighted breadth-first search algorithm implemented in *igraph* (25), we identify the geodesic (shortest path) connecting every pair of physicians in the patient-sharing network. We reduce geodesics to three categories: length 1, 2 or 3+ (where 3+ includes dyads that are not reachable in the network, i.e., whose physicians belong to separate components of the network).

We then calculate, for every pair of providers, the difference in each of our six specialty-adjusted measures of billing and prescribing intensity, and characterize the similarity of provider pairs as a function of their geodesic in the patient-sharing network. We use geodesics rather than binary connection status (connected vs. not connected) in our primary models of clustering in order to mitigate the possibility of clustering estimates being driven solely by the characteristics of a dyad's shared patients.

To determine the extent to which any observed clustering is produced by chance alone, we generate a "null distribution" of mean pairwise intensity differences by path length. For each of 1000 simulations, we retain the observed network structure but randomly permute, within each HRR, the intensity measures of providers. We then calculate, for each permutation, the mean pairwise difference in each of the intensity measures, by path length, and characterize the results of the simulations as a probability distribution. In such permutations, there is no association between the attributes of connected physicians.

Clustering of physician intensity: modeling approach. To control for unobserved attributes of the physicians comprising each dyad, we estimate a series of models, as follows. For each specialty-adjusted intensity measure (n=6), in each HRR (n=306), we model the dyadic difference in specialty-adjusted intensity of every pair of providers, as a function of an ordinal variable d_{ij} that encodes the length of the geodesic (1, 2, or 3+) separating the nodes i and j of the dyad, and random effects for both physicians party to the dyad. We estimate the following Bayesian model in *Stan* (26):

$$E\left[Y_{ij}\,\middle|\theta_i,\theta_j\right] = \alpha + \beta d_{ij} + \theta_i + \theta_j + \varepsilon_{ij},$$

where i and j denote the two physicians party to each dyad, Y_{ij} is the absolute difference of their specialty-adjusted intensity measures, d_{ij} is the length of the geodesic between i and j, and θ_i and ε_{ij} are provider- and dyad-specific latent factors related to the intensity difference between providers i and j, with $\theta_i \sim Normal(0,\tau^2)$, $\varepsilon_{ij} \sim Normal(0,\sigma^2)$, $\tau \sim cauchy(0,5)$,

$p(\alpha) \propto 1, p(\beta) \propto 1$ and $p(\sigma) \propto 1$. The distributions for the hyper-priors are diffuse in that they impart almost no information and, therefore, have minimal bearing on the analysis. The physician fixed effects θ_i and θ_j are notable for the involvement of two latent effects from the same structural level (physician) in the statistical model of each observation; here these latent physician effects account for unobserved attributes of the individual physicians comprising each dyad.

Approximation to Bayesian model. Estimation of the full Bayesian model is computationally demanding for large HRRs, due primarily to the large numbers of physician random effects and their additive specification for each observation. (The number of dyads *n* in any network grows with the square of the number of nodes *k*: $n = \frac{1}{2}k(k-1)$). For the 110 HRRs with fewer than 30,000 provider dyads (275 unique providers), we were able to estimate the full Bayesian model as well as a fast approximation (specifying separate fixed effects for each physician), using the method of Gaure (27, 28), which generalizes the *Frisch-Waugh-Lovell* theorem for projecting out categorical variables with large numbers of levels (29). This fast approximation replicates the point estimates of the Bayesian model, but for smaller HRRs produces falsely narrowed (anticonservative) confidence bounds (Fig. S1, left column), due to its incorrect modeling of physician i and physician j's fixed effects as separate distributions, i.e., permitting the same provider to have different fixed effects when they are coded as the "first" versus "second" provider in the dyad, which is meaningless in a non-directed network.

We therefore modify the method of Gaure (27, 28), extending the Frisch-Waugh-Lovell theorem and exploiting sparse matrix computations in *R* (30), to permit the rapid estimation of the effect of path length on dyadic intensity differences, correctly specifying additive physician fixed effects for each dyad, with a single fixed effect for each physician across all dyads to which she is party, regardless of her assuming position i or j in a given dyad.

For k physicians, jointly productive of $n = \frac{1}{2}k(k-1)$ dyadic observations of billing similarity and path length, the basic model in matrix form is as follows:

$$y = X\beta + D\alpha + \epsilon$$

where y is a vector of length n of pairwise (dyadic) differences in specialty-adjusted billing intensity, X is an $n \times (c+1)$ matrix of the geodesics separating each pair of providers and any c other dyadic covariates (such as provider concordance on experience or gender, q.v. below), D is an $n \times k$ matrix of dummy variables coding the physician fixed effects (constructed as detailed below), and ϵ is a normally distributed error term.

Performing ordinary least squares (OLS) regression on this system would yield estimates $\hat{\beta}$ for the effect of path length and any other dyadic covariates on intensity difference, and estimates $\hat{\alpha}$ for the physician fixed effects. For large HRRs, however, with tens of millions of observations and thousands of physician fixed effects, standard OLS estimation is computationally infeasible.

The Frisch-Waugh-Lovell theorem states that if P is the projection onto the orthogonal complement of the range of D, then the projected system

$$Py = PX\beta + P\epsilon$$

yields the same $\hat{\beta}$ when estimated with OLS. (The projection shares some of the mechanics of instrumental variables estimation in that the involvement of terms we wish to ignore (in the case of IV these are unmeasured confounders) are removed by an orthogonal projection.) The projected system does not contain the numerous physician fixed effects. Though P itself is $n \times n$, and thus computationally infeasible in the above system for large HRRs ($n \approx 10^7$), Py and Px are more easily computed using sparse matrix computations (30).

Thus, for the simplest of our models, in which the length of the shortest path (geodesic) separating each pair of physicians is the only dyadic predictor, we estimate as follows the effect of geodesic distance on differences in specialty-adjusted intensity:

1) For k physicians, jointly productive of $n = \frac{1}{2}k(k-1)$ dyadic observations of billing similarity and path length, create the sparse $n \times k$ matrix D, containing two 1's per row, indicating the two physicians party to each of the n dyads.

2) Project out the additive contributions of the physicians from each dyadic observation of path length, by computing a "physician residual" as follows:

$$PGeodesic = Geodesic - D\left((D \times D)^{-1}(D \times Geodesic)\right)$$

3) *Mutatis mutandis*, project out the additive contributions of the physicians from each dyadic observation of difference in billing intensity, as follows:

$$PDiff = Diff - D\left((D \times D)^{-1}(D \times Diff)\right)$$

4) Then, simply regress, with OLS, the dyadic intensity differences on geodesics, having projected out the physician fixed effects from each:

$$PDiff = \beta \cdot PGeodesic + \epsilon$$

5) The standard error of the estimate $\hat{\beta}$ must be corrected by adjusting the degrees of freedom according to the number of parameters projected out by P:

$$SE(\hat{\beta})_{corrected} = \sqrt{\frac{n}{(n-k)}} SE(\hat{\beta})_{naive}$$

This method precisely replicates both the point estimates as well as the confidence bounds of the properly specified Bayesian model (Fig. S1, right column). Due to its dramatic computational advantage over the latter (over two orders of magnitude difference in computing time), we use this method for all main analyses of clustering.

Models stratified by provider gender and experience. To investigate the effects of provider experience and gender on clustering of billing intensity, we also estimate models stratified by provider gender and experience. Provider gender is included in the data released by CMS. To estimate provider experience, we use the year the provider's NPI was enumerated into the National Plan and Provider Enumeration System (NPPES) database, labeling providers enumerated in 2007 or later as "new," and those enumerated prior to 2007 as "old" (Fig. S2). We then replicate the geodesic-based clustering models described above, separately for female-female, female-male, and male-male dyads, and for new-new, new-old, and old-old dyads.

Clustering of physician intensity: Assortativity. As an alternative measure of clustering, we calculate, for each HRR, physicians' assortativity on each measure of specialty-adjusted billing intensity. We use the method of Newman (31) for scalar characteristics.

We then compare assortativity coefficients to clustering estimates from the dyadic statistical models described above, calculating the Pearson correlation, for each measure, between the HRR-level estimates produced by the two methods.

Dyadic similarity by strength of edge. For directly connected ($d_{ij} = 1$) providers, we also characterize dyads' specialty-adjusted intensity differences Y_{ij} as a function of the strength of their relationships, where strength is defined by the edge weights defined above. We correlate Y_{ij} with each of the four measures of edge strength described above, using Spearman's *rho*.

Longitudinal analyses: non-parametric approach. To test for social influence among physicians as a potential mechanism contributing to observed clustering on intensity, we evaluate each dyad's *change* in Y_{ij} from 2012-2013, ΔY_{ij}, as a function of the dyad's change in connection status Δc_{ij} (i.e., remain unconnected, remain connected, disconnect, or connect). The rationale for this analysis is that social influence is only expected to act between pairs of physicians that are socially connected. Therefore, we expect Y_{ij} to decrease $\left(\Delta Y_{ij} < 0\right)$ after a tie develops between physicians *i and j*, to continue to converge if the tie is maintained, and to diverge $\left(\Delta Y_{ij} > 0\right)$ after the tie dissolves. However, because we do not compute change from the exact point at which their tie forms or dissolves, the effect for dyads that form a tie may capture part of the "selection" or homophily described below. Depending on the length of time that influence takes to act, and on whether influence converges to an equilibrium point, the effect for maintained ties and the effect for new ties may respectively under- and over-estimate the pure influence effect.

To test for selection as a mechanism of clustering, we evaluate each dyad's adjusted intensity differences in 2012, as a function of its change in connection status Δc_{ij}, as above. If clustering on intensity arises in part from physicians preferentially forming relationships with similar physicians, one would expect dyads that formed new connections in 2013 to have been more similar in 2012 than dyads that remained unconnected. Likewise, if physicians preferentially sever relationships with dissimilar physicians, one would expect dyads that severed connections in 2013 to have been *less* similar in 2012 than dyads that remained connected.

Longitudinal analyses: modeling approach. As with our cross-sectional analyses of clustering on physician intensity, we estimate models for influence and selection that account for the attributes of individual physicians. We employ the same estimation strategy described above, with models of the following form to evaluate for influence:

$$E\left[\Delta Y_{ij} \,\middle|\theta_i, \theta_j\right] = \alpha + \beta \Delta c_{ij} + \theta_i + \theta_j + \varepsilon_{ij},$$

where ΔY_{ij} is the change in dyadic intensity difference from 2012 to 2013, Δc_{ij} are indicators for the change in a dyad's connection status over this period (remain unconnected, remain connected, disconnect, or connect), and $\theta_i + \theta_j$ are fixed effects for each physician party to each dyad.

To evaluate for selection, we estimate models of the form:

$$E\left[nc_{ij} \,\middle|\theta_i, \theta_j\right] = \alpha + \beta Y_{ij(2012)} + \theta_i + \theta_j + \varepsilon_{ij},$$

where nc_{ij} is an indicator for the establishment of a new patient-sharing relationship in 2013 among previously unconnected providers, and $Y_{ij(2012)}$ is the dyad's specialty-adjusted intensity difference in 2012.

In supplemental models summarized in Table S3, we estimate the effect of tie dissolution (compared to tie maintenance) on change in specialty-adjusted difference:

$$E\left[\Delta Y_{ij} \,\middle|\theta_i, \theta_j\right] = \alpha + \beta sc_{ij} + \theta_i + \theta_j + \varepsilon_{ij},$$

where sc_{ij} is an indicator for tie disconnection from 2012 to 2013, and the other parameters are as above.

As a supplemental measure of selection (also summarized in Table S3), we estimate models of the form:

$$E\left[sc_{ij} \,\middle|\, \theta_i, \theta_j\right] = \alpha + \beta Y_{ij(2012)} + \theta_i + \theta_j + \varepsilon_{ij},$$

to determine the effect of a connected dyad's specialty-adjusted intensity difference in 2012 on the probability of its severing its connection in 2013.

It may appear surprising that we use linear regression to model the probability of connection formation and dissolution. A weakness of the so-called "linear probability model" is that it does not constrain fitted values to the interval [0, 1] and does not automatically account for the fact that the variance of observations depends on the mean, with much smaller variability at 0 and 1 than for means (probabilities) near 0.5. These two concerns manifest as a loss of statistical efficiency compared to fitting a binary regression model such as logistic regression. However, given the scale of data we have available, statistical efficiency is of minor concern compared to possible bias from not accounting for the additive physician fixed-effect structure of the regression equation, as would be the case if we either ignored physician effects altogether or allowed separate effects depending on whether a physician corresponded to the first (*i*) or second (*j*) physician in a dyadic observation. The use of the linear probability model allows the use of Gaure's approximation to estimate our model using standard software on the full data set, with appropriately specified physician latent factors.

For the supplemental "disconnection" models described in Table S3, we estimate a single model for each intensity measure (rather than aggregating 306 HRR-level estimates), and use the unmodified method of Gaure (27), with "independent" physician fixed effects, for computational reasons. Thus, the confidence intervals for these supplemental models only may be anticonservative to the extent depicted in the left column of Fig. S1.

Results

In 2013, there were 2,533,452 unique, undirected pairs of connected providers, with both providers practicing a referral-relevant specialty and billing Medicare for at least 100 outpatient office visits. Eligible provider dyads shared a median of 35 patients.

Though routine office visits are represented by only five of the 5,983 service codes for which physicians billed Medicare Part B (the arm of Medicare covering outpatient services) in 2013, they account for 18.1% of the $100.2 billion paid to Medicare Part B providers for services that year (Fig. 1), and providers who bill for established patient office visits receive a median of 47.3% of their Medicare Part B income from such visits. Most visits are billed at medium intensity, but the distribution of visit intensities varies widely by provider, with 7.9% of providers billing exclusively at the highest intensity levels, 4 and 5 (Fig. S3). Billing intensity also varies regionally: in some HRRs, no providers bill exclusively at the highest levels, while in others fully 17.7% of providers do so.

In 2013, 94.2% of dyads were mutually reachable in the patient-sharing network by geodesics of 1-18 steps. 2.0% of pairs were directly connected, i.e., shared 11 or more patients within a 30-day window. 13.4% of provider pairs were on geodesics of length 2, i.e., did not themselves share 11 or more patients, but each shared 11 or more patients with a mutual colleague. 84.6% of pairs were separated by geodesics of length 3 or more (Fig. S4).

78.0% of HRR-level clustering estimates β (i.e., estimates for the effect of geodesic on intensity difference) are individually positive and significant ($p < 0.05$, Fig. 2G): controlling for the attributes of individual physicians, pairs of physicians who are connected by shorter paths in the patient-sharing networks are more similar on all specialty-adjusted measures of clinical, billing and prescribing intensity. We used meta-analytic methods to summarize the estimates for each HRR in the form of nationwide clustering estimates (Fig. 2G), all of which are positive and significant ($p < 0.001$). This clustering suggests that aberrant billing behavior is not merely the result of isolated “bad actors”, but rather has widespread, systemic causes.

These model-based estimates are highly compatible with the results of two nonparametric approaches, based on assortativity (31)(Fig. S5), and random permutations of provider attributes (Table S1, Fig. S6). For 5 of 6 measures, HRR networks exhibiting greater clustering are both more extreme and more varied in the intensity of their physicians (Fig. S7). Among pairs of connected (d_{ij}=1) providers, ties constituted by a greater proportion of each provider's total shared patients and services exhibit greater similarity in specialty-adjusted billing intensity than ties constituted by fewer of the providers' total patients and services (Table S2).

We also investigated how physicians' clustering on intensity might vary by clinical experience ("new" for inexperienced physicians, "old" for experienced physicians) and sex (male or female). In models stratified by the estimated practice experience of the physicians party to each dyad, new-new dyads exhibit significantly greater evidence of clustering on all intensity measures than old-new or old-old dyads. Stratifying by provider sex, female-female dyads exhibit greater clustering than male-female or male-male dyads (Fig. S8).

Figure 3 depicts the wide geographic variation in both the proportion of high-intensity providers (a *nodal* attribute), as well as in the estimated degree of clustering on this attribute (a *structural*, network-level measure). Fig. 4 illustrates the nature of clustering for two pairs of HRRs with similar numbers of total and high-intensity providers, but with different degrees of provider *clustering* on intensity.[1]

In a social network, clustering on any mutable attribute, such as physicians' billing and prescribing behavior, can result from any combination of *influence* (connected physicians induce similarity in one another, e.g., by imitation or persuasion), *selection* (physicians choose to establish or maintain relationships with already similar providers), or *context* (connected physicians share a common influence that affects them both). In additional analyses, we exploit the longitudinal nature of the data to investigate separately for influence and selection in the production of the clustering we observe (Figs. 2 and S9).

[1] Interactive maps of billing and network attributes, and representative networks exhibiting low and high degrees of clustering are available online.

First, to investigate the potential effect of social influence, we estimate the effect of change in connection status Δc_{ij}, i.e., whether a dyad is maintained (remains connected), formed, or dissolved, on the change in Y_{ij} for all dyads, between 2012 and 2013. Because influence is expected to propagate only between connected physicians, we would expect, under social influence, for Y_{ij} to decrease following tie formation (Fig. 2D), to continue to converge following tie maintenance (Fig. 2C), and to diverge following tie dissolution. For four measures of intensity, dyads that form new connections become significantly more alike than dyads remaining unconnected. For three measures, dyads that remain connected grow more similar than those remaining unconnected (Fig. 2H). For one measure, dyads that sever connections grow significantly less alike than those that remain connected, while for the other measures, no significant effect is observed (Table S3). These results support influence as one mechanism of clustering: connected providers induce one another to adopt similar behaviors.

Second, to investigate the potential effect of selection, we estimate the effect of $Y_{ij(2012)}$ between unconnected providers on the probability of their forming a connection in 2013 (Fig. 2E), and the effect of $Y_{ij(2012)}$ between connected providers on the probability of their dissolving their connection in 2013 (Fig. 2F). We find, for all intensity measures, that greater similarity (smaller Y_{ij}) among unconnected dyads in 2012 is associated with a greater predicted probability of forming a connection in 2013 (Fig. 2I), and that greater dissimilarity (larger Y_{ij}) among connected dyads is associated with a greater predicted probability of dissolving their connection the following year (Table S3). These findings are consistent with selection, whereby physicians preferentially establish or maintain relationships with similar providers. Nonparametric methods support the same substantive results (Fig. S9).

Discussion

Physician behavior drives the quality and costs of health care (11). Because physicians have wide discretion in choosing which services to perform in a given clinical situation, which drugs to prescribe, and even in how to bill for their work, physicians of the same specialty differ widely in the intensity of care furnished for a given patient, with major implications for the quality, cost and equitable distribution of health care nationwide (8, 10, 32-34).

Using multiple nonparametric approaches as well as dyadic models estimating pairwise similarity as a function of geodesic and physician attributes, we find that physicians who share patients are more similar on six dimensions of billing intensity than physicians who do not share patients but share a common colleague, who in turn are more similar than pairs of physicians separated in the network by three degrees or more. In longitudinal models, we find evidence for both influence and selection in the production of the clustering we observe (Fig. 2).

Although influence is notoriously difficult to distinguish from selection in observational network data (35-37), and cannot be definitively established from these data, converging lines of evidence suggest a role of social influence in producing the clustering we observe. Billing and prescribing intensity are learned in clinical practice, grant wide discretion to the individual physician, and are weakly constrained by objective guidelines; all qualities rendering them ripe for peer influence. Patient-sharing ties derived from administrative data are strongly associated with relationships reported by physicians themselves to be channels of informal clinical advice (38), and physicians who move to more or less aggressive regions rapidly update their own behavior to better approximate the local norm (15). Young physicians learn primarily by watching others (6), and indeed, we find significantly greater clustering among new physicians, as would be expected if clustering on intensity arose in part from imitation. Likewise, some evidence suggests that women weigh perceived social norms more heavily than men in their decisions to adopt technologies and behaviors (39), and indeed we find significantly greater clustering among female-female than among male-male or mixed-gender dyads.

Our study has limitations. With only two years of publicly available billing data, the relative contributions of influence and selection to the clustering we observe cannot be precisely resolved. The billing data themselves, while unprecedented in scope and completeness, are aggregated by service, drug and provider, and the appropriateness of any individual service cannot be assessed against a particular clinical episode. (When such audits have been done, however, an enormous proportion of routine office visits were found to be billed at levels inconsistent with the clinical record (40).) Because our intensity measures are averaged across at least 100 such episodes for every provider, our focus is less on instances of individual fraud (for which different mechanisms may be responsible) than on physicians' *styles* of clinical and billing intensity, and on the social patterning of those styles. Nevertheless, aggregated data do permit the detection of aberrant behavior: providers billing exclusively at the highest intensity levels for hundreds of routine visits, for instance, are exceedingly unlikely to have billed in strict accordance with clinical reality.

That physicians cluster systematically on the basis of their clinical and billing intensity, even after controlling for the attributes of individual providers, is of direct relevance to health policy and economics. Physicians and patients are embedded in complex networks of formal and informal influence, and many current proposals to improve the value and quality of health care, such as the Accountable Care Organization, pertain directly to the structure and function of provider networks (41, 42). Policy interventions designed to alter physician behavior are likely to be more effective when tailored to the networks through which they are intended to diffuse (43). In particular, the optimal design and delivery of behavioral interventions is likely to vary by the degree and pattern of clustering in physician networks (44). The pairs of HRR networks depicted in Fig. 4, for instance, have similar numbers of physicians and similar proportions of high-intensity providers, but the distribution and clustering of those providers, and the patterns of patient-sharing among them, vary dramatically. Regardless of whether clustering arises from selection or influence, efforts to reduce inappropriate billing may benefit from targeting not only known high-intensity providers, but also providers newly associated with high-intensity clusters, who may be susceptible to adopting or reflecting the local norm (15). In other cases, policies might be designed to alter the structure of the network itself, for instance by incenting physicians to redirect referrals from low- to high-value providers (45).

The comprehensive scope of newly available public data, such as the billing, prescribing and patient-sharing records we study here, permit novel approaches to problems of broad societal interest. But the significance of such data exceeds the individual studies to which the data are applied (1). In our view, calls for greater transparency in economic and public affairs, however laudable, are too often restricted to the identification of individual "bad actors." Social actors and their actions, however, are graded and dynamic, and are embedded in social systems that can dampen or diffuse the behaviors and ideas with which they are seeded. By democratizing access to detailed, individual-level information about the allocation of public funds, the movement toward open data will naturally encourage scrutiny of the actors who embody the complex systems in whose function virtually everyone bears a stake. By restoring these individual data to the relational structures under which they arose, researchers can encourage the illumination of those systems, as well.

Summary

In 1973, Wennberg and Gittelsohn launched an enormously influential line of research on geographic variations in the costs and quality of health care. This has since become the paramount empirical argument for the inefficiency and waste of the American health care system. But this line of research has tacitly treated geographic variation in health care intensity, costs and quality as though they were *bona fide* geographic phenomena, such as weather patterns, or terrain. In reality, because health care costs and quality are driven primarily by physician behavior, these variations likely reflect more fundamental *social* phenomena: we have simply lacked the data and methods to resolve health care variations to this level. 43 years after the publication of Wennberg and Gittelsohn's seminal paper, open-access data, computational capacity, and suitable statistical approaches enable us to address this problem, and to rigorously characterize the "social geography" of American medicine.

Scientific authorities have recently drawn needed attention to the crisis of reproducibility in scientific research. In our view, the need for open data and readily reproducible methods is greater still for research on sensitive and socially contentious topics such as the annual misuse of billions of taxpayer dollars in health care. A recent Health and Human Services audit found that fully 42% of Medicare claims for evaluation and management services in 2010 were coded incorrectly, while another recent investigation found hundreds of millions of dollars of fraudulent charges to Medicare's prescription drug program. Such reports draw attention to the enormous inefficiency, waste and fraud plaguing the health care system, but the potential to translate these findings into specific policy and research agendas is limited by the inaccessibility of the data and methods used, the several-year lag between data generation and report publication, and the inability to integrate the findings with other relevant datasets to address questions not specifically included in the report.

In this paper we use very recent, comprehensive, and publicly available data, and new analytic methods, to probe the social patterning of physician behavior, with implications for patients, taxpayers, and health reform at large. In April 2015, The Centers for Medicare & Medicaid Services (CMS) publicly released, for the first time, comprehensive data on all prescriptions

billed to Medicare Part D in 2013. In June 2015, CMS released detailed service and payment data for the millions of providers billing Medicare Part B in 2013, as well as updated data for 2012. Finally, in December 2015, CMS released comprehensive patient-sharing data for Medicare providers between 2009 and 2015.

This paper represents both a new synthesis of these unprecedented health care data (to our knowledge, involving the largest and most complete network of physicians constructed to date), as well as a proof-of-principle for a new school of computational social science that is more open, transparent, reproducible, responsive and relevant to public affairs. By characterizing the dynamic patient-sharing networks that both shape and reflect the behavior of physicians, we aim to elucidate the social micro-foundations underlying the enormous cost and complexity of American health care, and to open new avenues for data-driven policy intervention. By making our data, code and methods readily available, we aim to create a framework for integrating and analyzing all of the provider-level data thus far made public by CMS, and invite other researchers to expand upon and refine our methods, and to pose questions beyond those addressed here. We also create interactive visualizations to encourage non-experts to meaningfully explore the data, to gain intuition for the social processes underlying health care, and to develop their own questions about health care and physician behavior.

The comprehensive scope of newly available public data, such as the billing, prescribing and patient-sharing records we study here, permit novel approaches to problems of broad societal interest. But the significance of such data exceeds the individual studies to which the data are applied. In our view, calls for greater transparency in economic and public affairs, however laudable, are too often restricted to the identification of individual "bad actors." Social actors and their actions, however, are graded and dynamic, and are embedded in social systems that can dampen or diffuse the behaviors and ideas with which they are seeded. By democratizing access to detailed, individual-level information about the allocation of public funds, the movement toward open data will naturally encourage scrutiny of the actors who embody the complex systems in whose function virtually everyone bears a stake. By restoring these individual data to the relational structures under which they arose, researchers can encourage the illumination of those systems, as well.

References

1. Martin EG, Helbig N, Shah NR. Liberating data to transform health care: New York's open data experience. JAMA. 2014;311(24):2481-2. doi: 10.1001/jama.2014.4610.
2. Schneeweiss S. Learning from Big Health Care Data. New England Journal of Medicine. 2014;370(23):2161-3. doi: doi:10.1056/NEJMp1401111. PubMed PMID: 24897079.
3. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014;311(24). doi: 10.1001/jama.2014.4228.
4. Keehan SP, Cuckler GA, Sisko AM, Madison AJ, Smith SD, Stone DA, et al. National Health Expenditure Projections, 2014–24: Spending Growth Faster Than Recent Trends. Health Affairs. 2015;34(8):1407-17. doi: 10.1377/hlthaff.2015.0600.
5. Chernew ME, Newhouse JP. Health Care Spending Growth. In: Pauly MV, Mcguire TG, Barros PP, editors. Handbook of Health Economics. 2: Elsevier; 2011. p. 1-43.
6. Fuchs VR, Milstein A. The $640 Billion Question — Why Does Cost-Effective Care Diffuse So Slowly? New England Journal of Medicine. 2011;364(21):1985-7. doi: doi:10.1056/NEJMp1104675. PubMed PMID: 21591938.
7. Wennberg JE, Gittelsohn A. Small Area Variations in Health Care Delivery. Science. 1973;182(4117):1102-8.
8. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder ÉL. The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. Annals of Internal Medicine. 2003;138(4):273-87.
9. Baicker K, Chandra A. Medicare Spending, The Physician Workforce, And Beneficiaries' Quality Of Care. Health Affairs. 2004;W4:184-97.
10. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder ÉL. The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. Annals of Internal Medicine. 2003;138(4):288-98.
11. Cutler D, Skinner J, Stern AD, Wennberg D. Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending. National Bureau of Economic Research Working Paper Series. 2013;No. 19320. doi: 10.3386/w19320.
12. Milstein A, Gilbertson E. American Medical Home Runs. Health Affairs. 2009;28(5):1317-26. doi: 10.1377/hlthaff.28.5.1317.
13. Chandra A, Skinner J. Technology growth and expenditure growth in health care. NBER working paper. 2011.
14. Fisher ES, Bynum JP, Skinner JS. Slowing the Growth of Health Care Costs — Lessons from Regional Variation. New England Journal of Medicine. 2009;360(9):849-52. doi: doi:10.1056/NEJMp0809794. PubMed PMID: 19246356.
15. Molitor D. The evolution of physician practice styles. Dissertation. 2014.
16. Landon BE, Keating NL, Barnett ML, Onnela J-P, Paul S, O'Malley AJ, et al. Variation in patient-sharing networks of physicians across the united states. JAMA. 2012;308(3):265-73. doi: 10.1001/jama.2012.7615.
17. Centers for Medicare & Medicaid Services. Medicare Provider Utilization and Payment Data: Physician and Other Supplier. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier.html. 2015.

18. Centers for Medicare & Medicaid Services. CMS releases prescriber-level Medicare data for first time. http://cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-04-30.html. 2015.
19. ProPublica. Medicare Part D Prescribing Data 2012. https://projects.propublica.org/data-store/sets/health-mcd12-1. 2014.
20. Dartmouth Altas of Health Care Working Group. The Dartmouth Atlas of Health Care. http://www.dartmouthatlas.org. 2015.
21. Newman MEJ. Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E. 2001;64(1):016131.
22. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E. 2001;64(1):016132.
23. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(11):3747-52. doi: 10.1073/pnas.0400087101.
24. Dartmouth Altas of Health Care Working Group. The Dartmouth Atlas of Health Care in the United States. Wennberg JE, editor: American Hospital Publishing, Inc.; 1996.
25. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695.
26. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Version 26. 2015;http://mc-stan.org/.
27. Gaure S. OLS with multiple high dimensional category variables. Computational Statistics & Data Analysis. 2013;66(0):8-18. doi: http://dx.doi.org/10.1016/j.csda.2013.03.024.
28. Gaure S. lfe: Linear Group Fixed Effects. The R Journal. 2013;5(2).
29. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge: MIT Press; 2010.
30. Bates D, Maechler M. Matrix: Sparse and Dense Matrix Classes and Methods. R package version 12-0. 2015.
31. Newman MEJ. Mixing patterns in networks. Physical Review E. 2003;67(2):026126.
32. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional Variations in Diagnostic Practices. New England Journal of Medicine. 2010;363(1):45-53. doi: doi:10.1056/NEJMsa0910881.
33. Skinner J. Causes and Consequences of Regional Variations in Health Care. In: Pauly MV, Mcguire TG, Barros PP, editors. Handbook of Health Economics. 2: Elsevier; 2011. p. 45-93.
34. Baicker K, Buckles KS, Chandra A. Geographic Variation In The Appropriate Use Of Cesarean Delivery. Health Affairs. 2006;25(5):w355-w67.
35. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci USA. 2009;106(51):21544-9. doi: 10.1073/pnas.0908800106.
36. VanderWeele TJ. Sensitivity Analysis for Contagion Effects in Social Networks. Sociological Methods & Research. 2011;40(2):240-55. doi: 10.1177/0049124111404821.
37. Shalizi CR, Thomas AC. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociological Methods & Research. 2011;40(2):211-39. doi: 10.1177/0049124111404820.

38. Barnett ML, Landon BE, O'Malley AJ, Keating NL, Christakis NA. Mapping Physician Networks with Self-Reported and Administrative Data. Health Services Research. 2011;46(5):1592-609.
39. Venkatesh V, Morris MG. Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. MIS Quarterly. 2000;24(1):115-39. doi: 10.2307/3250981.
40. Department of Health and Human Services. Improper payments for evaluation and management services cost Medicare billions in 2010. OEI-04-10-00181. 2014.
41. Fisher ES, McClellan MB, Safran DG. Building the Path to Accountable Care. New England Journal of Medicine. 2011;365(26):2445-7. doi: doi:10.1056/NEJMp1112442. PubMed PMID: 22204720.
42. Landon BE, Onnela J-P, Keating NL, Barnett ML, Paul S, O'Malley AJ, et al. Using Administrative Data to Identify Naturally Occurring Networks of Physicians. Medical Care. 2013;51(8):715-21.
43. Valente TW. Network Interventions. Science. 2012;337(6090):49-53. doi: 10.1126/science.1217330.
44. Centola D. An Experimental Study of Homophily in the Adoption of Health Behavior. Science. 2011;334(6060):1269-72. doi: 10.1126/science.1207055.
45. Song Z, Sequist TD, Barnett ML. Patient referrals: A linchpin for increasing the value of care. JAMA. 2014. doi: 10.1001/jama.2014.7878.

Figures and Tables

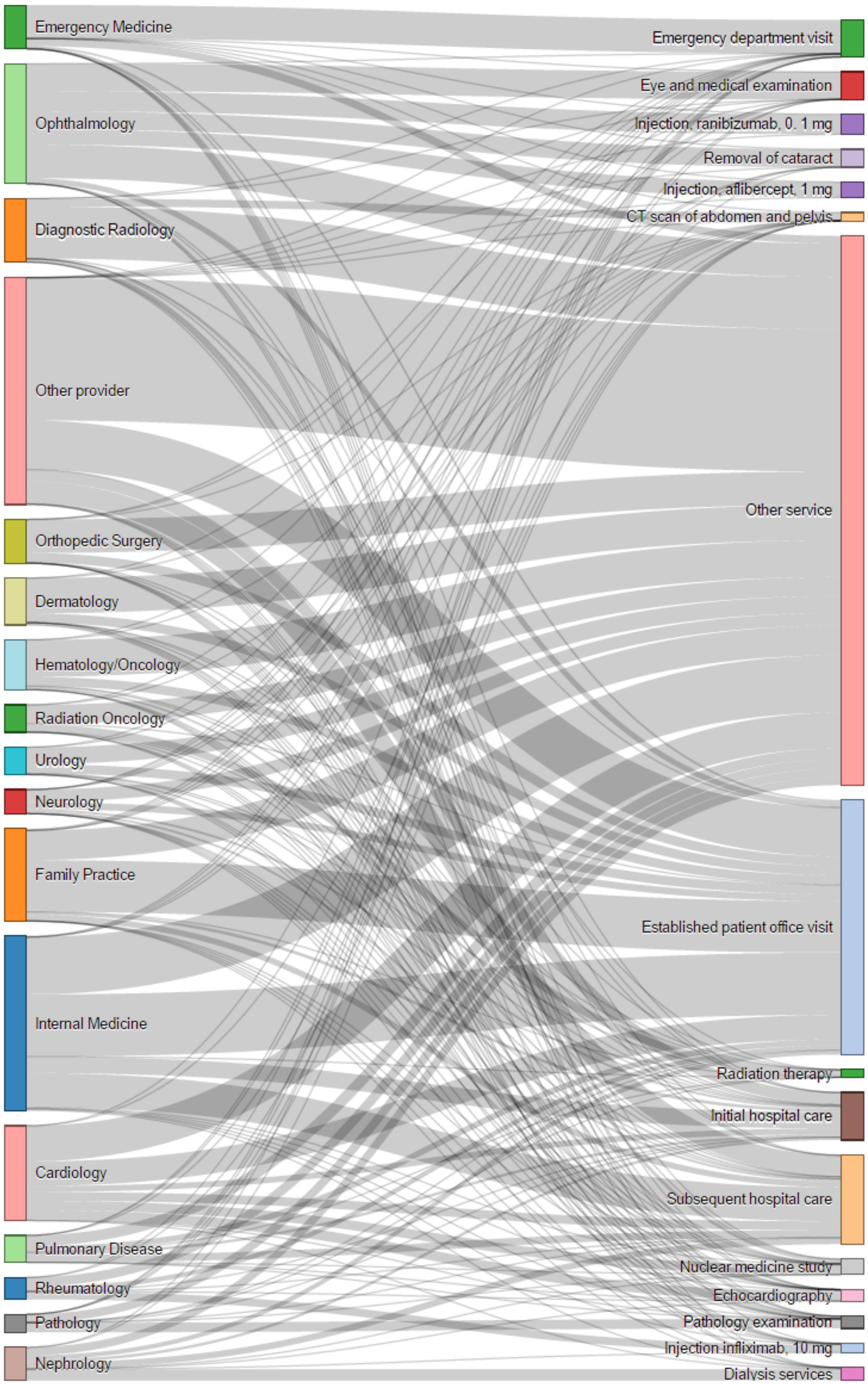

Fig. 1. Relationships between specialties and services in Medicare Part B. A bipartite graph depicts the allocation of Medicare Part B spending in 2013 among physician specialties and services. Height on the vertical axis is proportional to spending. Established patient office visits account for 18.1% of the $100.2 billion paid to Medicare Part B providers in 2013.

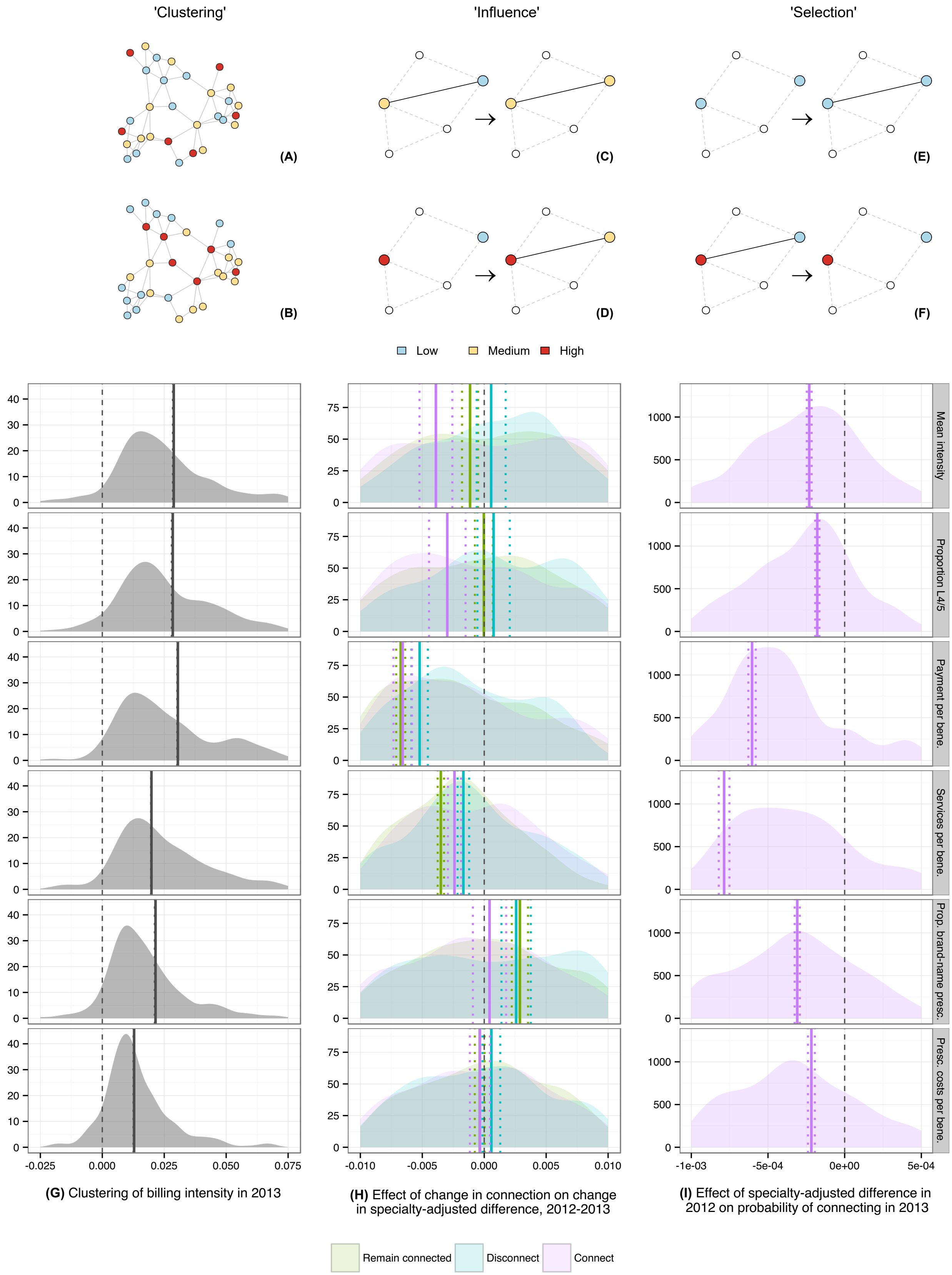
'Clustering'
'Influence'
'Selection'
(A)
(B)
(C)
(D)
(E)
(F)
Low
Medium
High
Mean intensity
Proportion L4/5
Payment per bene.
Services per bene.
Prop. brand-name presc.
Presc. costs per bene.
(G) Clustering of billing intensity in 2013
(H) Effect of change in connection on change in specialty-adjusted difference, 2012-2013
(I) Effect of specialty-adjusted difference in 2012 on probability of connecting in 2013
Remain connected
Disconnect
Connect

Fig. 2. Clustering of physician intensity can arise from both influence and selection.

(A) and **(B)** depict a small, simulated physician network with billing intensity levels (specialty-adjusted proportion of office visits billed at levels 4 or 5) drawn from the empirical distribution (here, physician intensity is "low" if below the 40th percentile and "high" if above the 80th). In **(A)**, dyads' geodesic separation d_{ij} in the network is unrelated to their specialty-adjusted intensity difference Y_{ij} (the clustering coefficient β is 0). In **(B)**, the network structure and distribution of intensities is the same as in **(A)**, but dyads more proximate in the patient-sharing network (lower d_{ij}) are more similar in billing intensity ($\beta =$ 0.10). **(C)** and **(D)** demonstrate how clustering can arise from *influence* among physicians. In **(C)**, the two highlighted physicians are dissimilar in intensity in 2012. In 2013, they maintain their patient-sharing relationship, and their pairwise intensity difference decreases ($\Delta Y_{ij} < 0$). In **(D)**, two dissimilar physicians establish a patient-sharing relationship, and again $\Delta Y_{ij} < 0$. Under *selection*, pairs of unconnected physicians preferentially form connections if they were already similar at baseline (low Y_{ij}, **(E)**), or terminate relationships with dissimilar physicians (high Y_{ij}, **(F)**).

In **(G)-(H)**, density plots depict the distribution of corresponding model estimates for each of the United States' 306 HRRs. Solid vertical lines denote pooled nationwide estimates (from fixed-effects meta-analysis of HRR-level results, i.e., larger HRRs weighted more heavily). Corresponding dashed lines indicate 95% confidence intervals. **(G)** depicts estimates of clustering on intensity in 2013, in which all measures show positive evidence of clustering ($p < 0.001$): controlling for the attributes of individual physicians, directly connected providers are more alike than pairs of providers separated by a shared colleague, who in turn are more alike than pairs of providers on geodesics of length 3 or more. **(H)** shows estimates of the effect of dyads' change in connection status on ΔY_{ij}: for the first four measures, pairs of physicians who form connections between 2012 and 2013 become significantly more alike than pairs who remain unconnected. **(I)** shows estimates of the effect of Y_{ij} in 2012 on the probability of forming a new connection in 2013: for all intensity measures, greater pairwise similarity among unconnected pairs in 2012 (lower Y_{ij}) predicts a higher probability of forming a connection in 2013.

Collectively, these results demonstrate robust social clustering of physician behavior, to which both influence and selection among physicians may contribute.

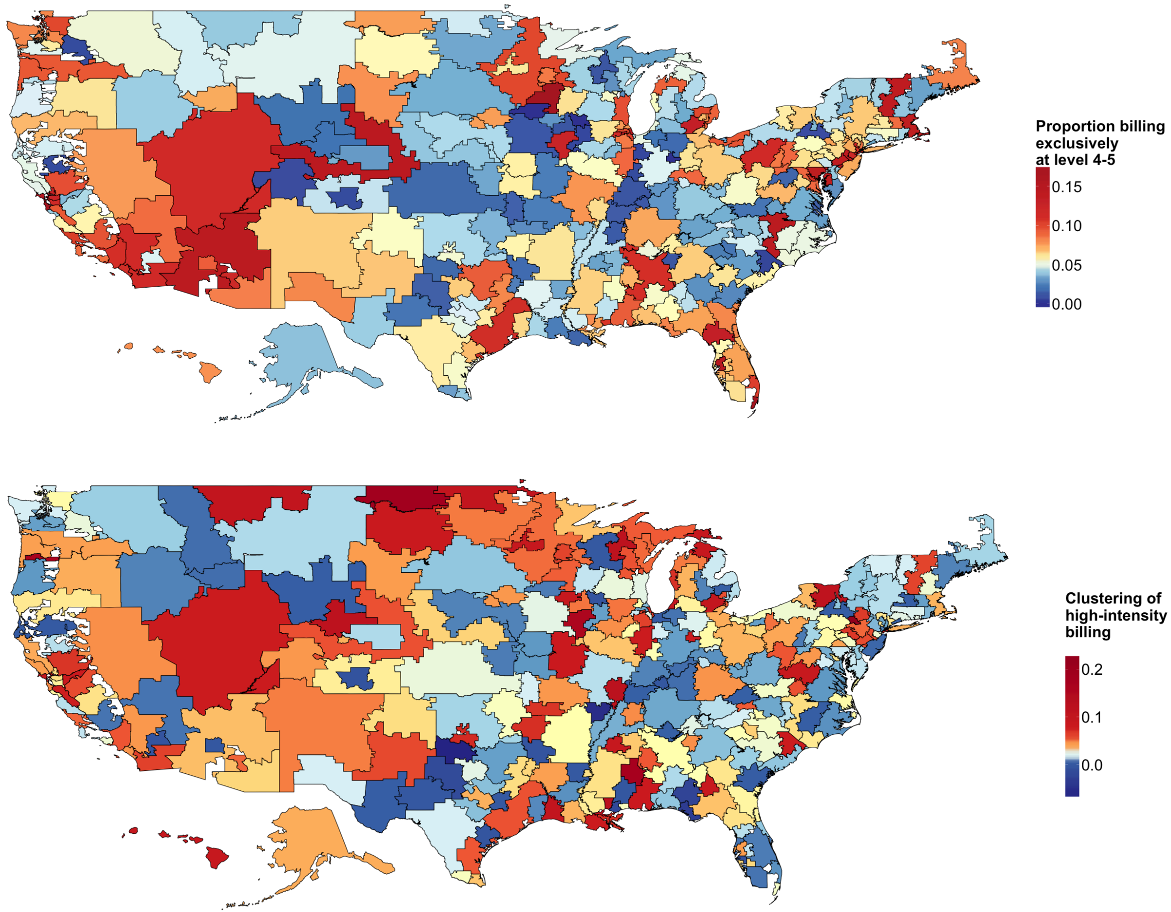

Fig. 3. Distribution and clustering of high-intensity billing. Traditional measures of geographic variation in health care have focused on variation in *nodal* means, such as the proportion of high-intensity providers in an HRR (upper panel). Combining billing and patient-sharing data permits analysis of geographic variation in *structural* features of health care, such as the clustering of billing practices in local physician networks (lower panel).

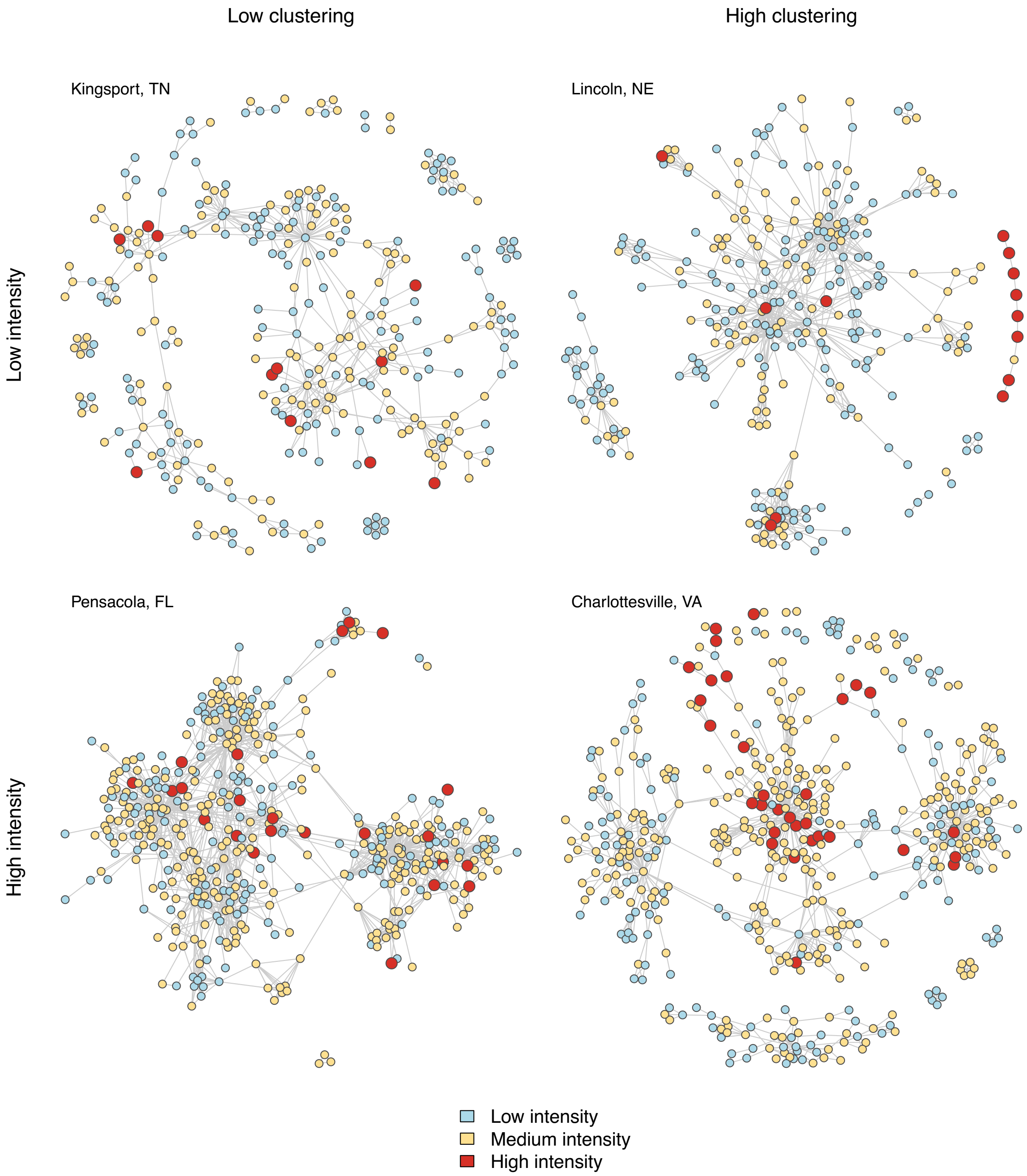

Fig. 4. Clustering of billing intensity in four health referral regions. Kingsport, TN (n=364 providers of eligible specialties billing at least 100 established patient office visits in 2013) and Lincoln, NE (n=331) each have relatively few providers billing exclusively at the highest intensity levels (3.6% and 3.9%, respectively), while Pensacola, FL (n=487) and Charlottesville, VA (n=566) have 6.8% and 6.9%. The clustering of billing intensity, however, differs markedly within each pair: in Lincoln and Charlottesville, physicians are more likely to exhibit patient-sharing relationships with providers of similar intensity. For legibility, only the top 20% of patient-sharing ties are shown, where ties are weighted by the provider-adjusted volume of patients shared. Physicians are labeled as "high" intensity if 100% of their office visits are billed at levels 4-5, and "low" if less than 50% of their visits are billed at these levels.

Supplementary Figure and Tables

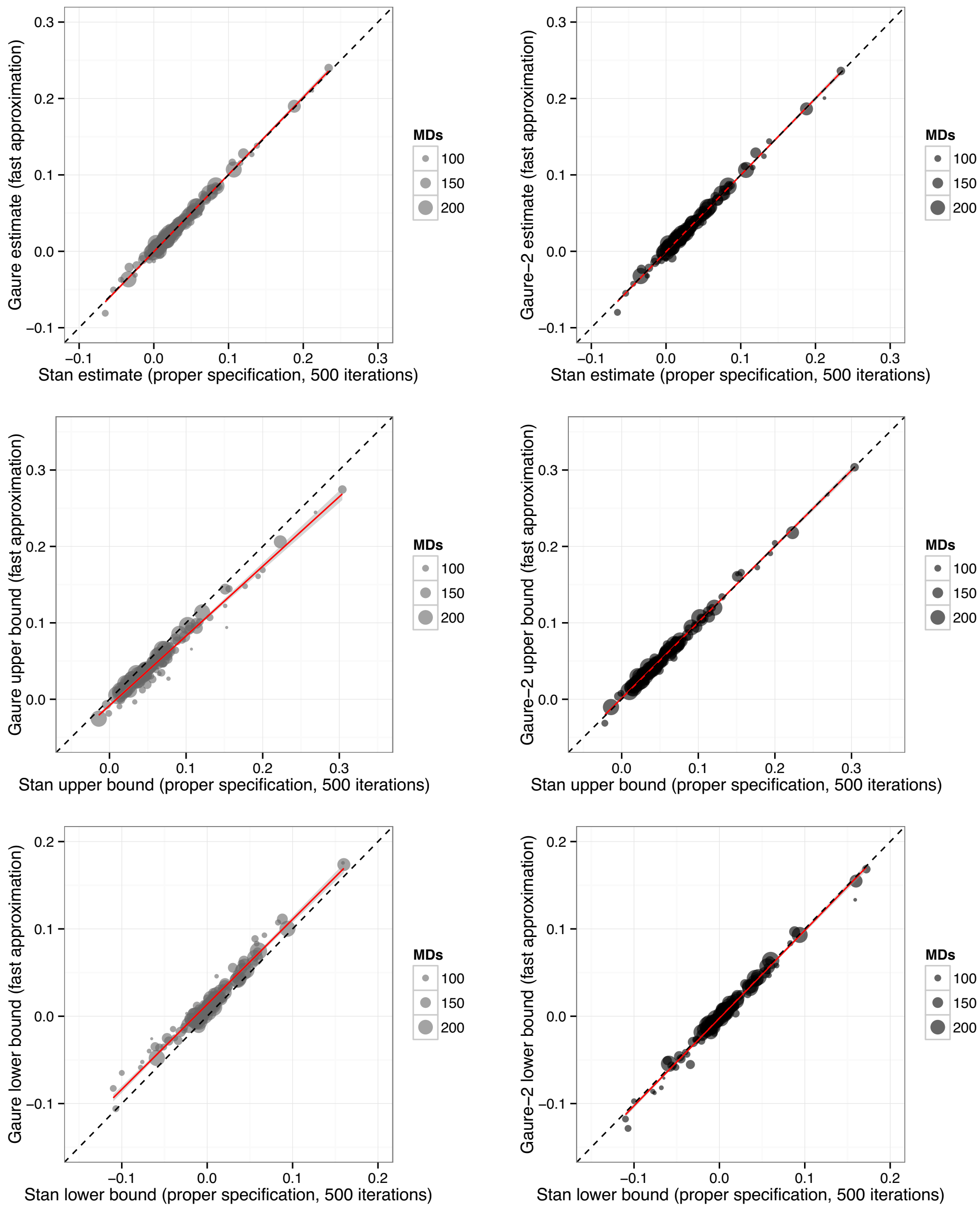

Fig. S1. Comparison between full Bayesian estimation and two approximations. Estimation of the clustering models is computationally intensive for large HRRs. The left column compares, for the 110 HRRs with fewer than 30,000 physician pairs, the results of a fully Bayesian estimation of the clustering model in Stan to a fast approximation using the method of Gaure: the fast approximation replicates the point estimates of the Bayesian model, but confidence bounds are falsely narrow due to the improper specification of physician fixed effects. The right column depicts the same comparison, using a modification of Gaure's (27, 28) method that permits the additive specification of physician fixed effects for each dyad: this method precisely replicates both the point estimates as well as the confidence bounds of the properly specified Bayesian model.

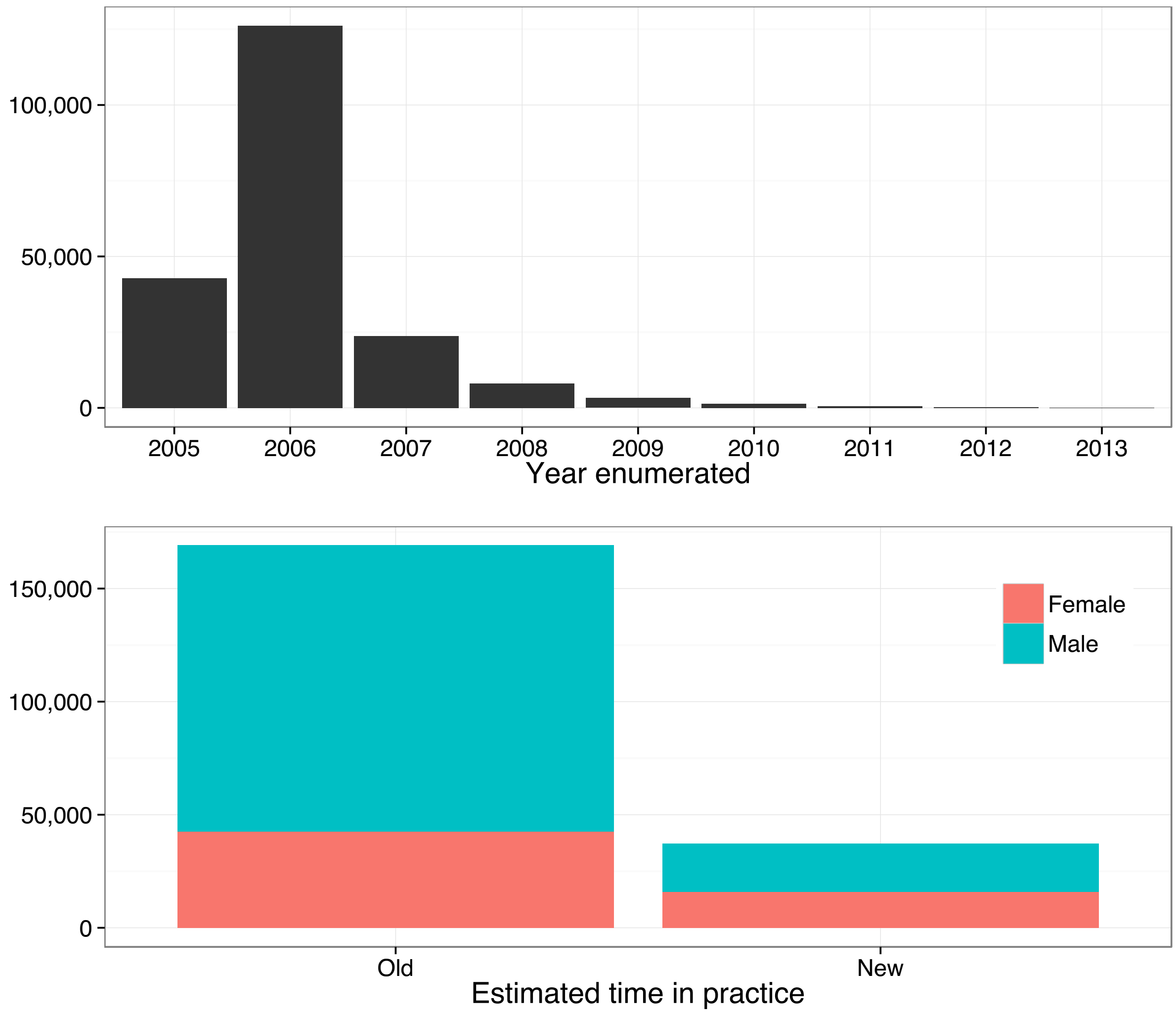

Fig. S2. Estimation of provider experience. The NPPES database was created in 2005, and most extant providers were enumerated by 2006 (upper panel). We classify providers as "old" if enumerated prior to 2007, and "new" if enumerated in 2007 or later. Old providers are predominantly male, while new providers are male and female in approximately equal proportion (lower panel).

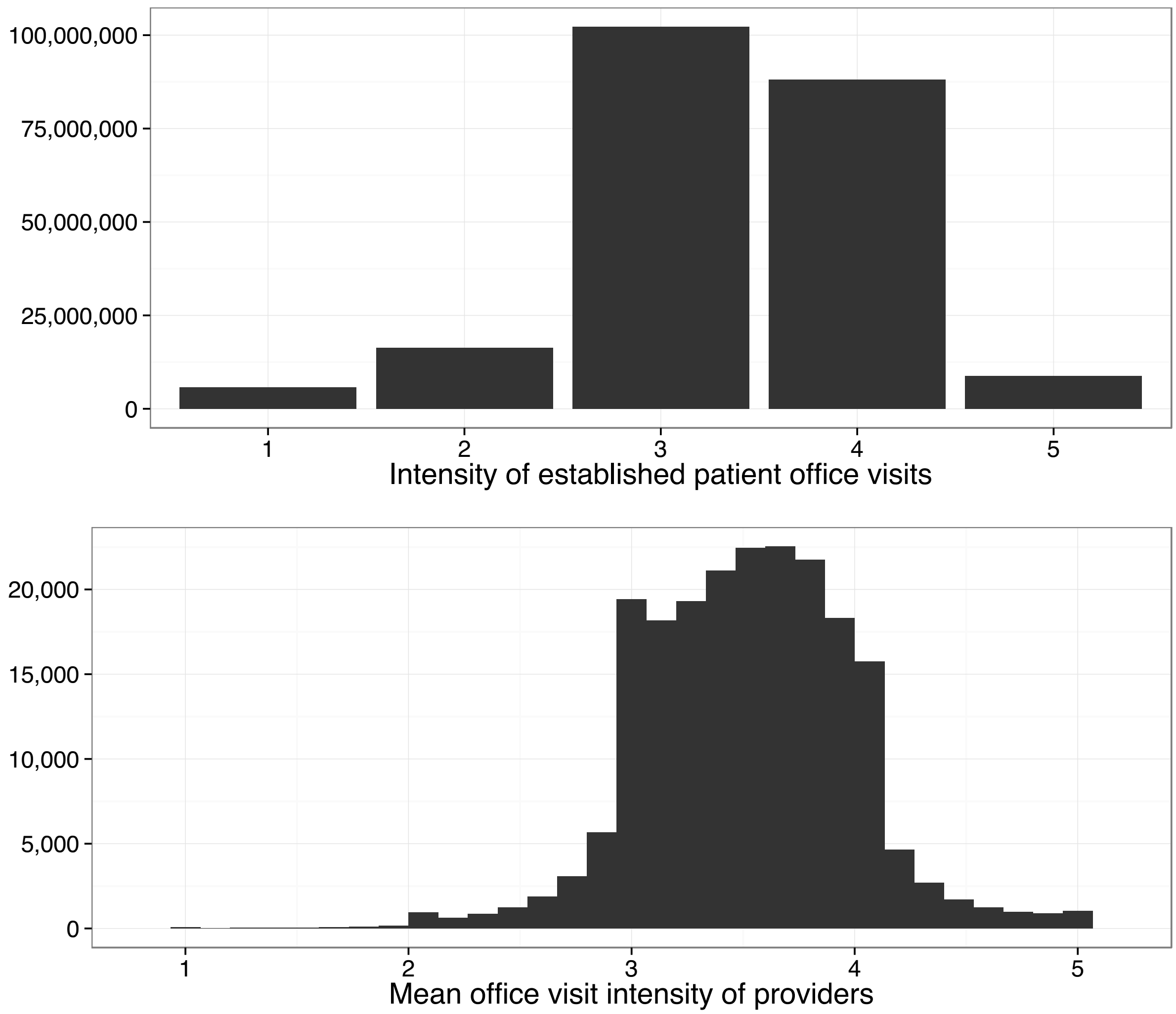

Fig. S3. Intensity of established patient office visits. In 2013, 469,853 Medicare providers were reimbursed for 221,322,530 office visits with established patients (on average, each eligible provider billed for 201 such visits). The upper panel shows that 86.0% of visits were billed at levels 3-4. The lower panel, however, which shows the mean billing level of the 206,873 providers of included specialties billing at least 100 visits in 2013, demonstrates that some providers billed much more intensely than others, with 7.9% of these providers billing exclusively at the highest intensity levels (4 and 5).

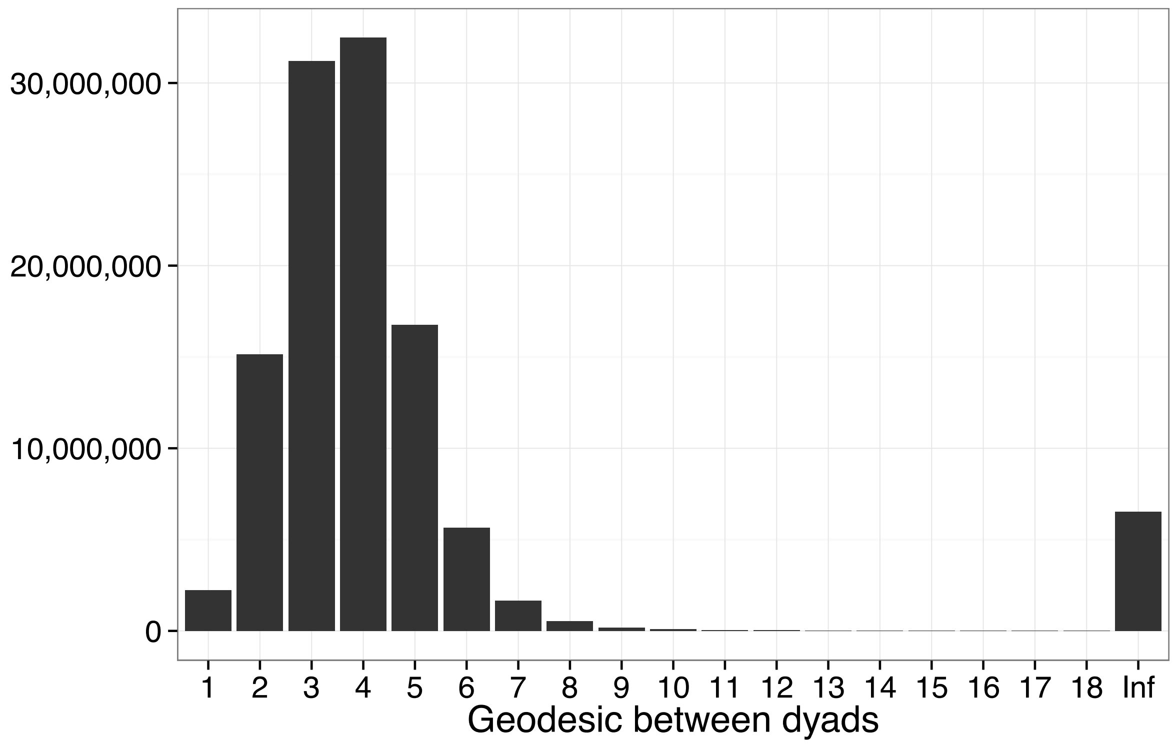

Fig. S4. Shortest path lengths (geodesics) between Medicare providers nationwide. In 2013, there were 112,547,233 pairs of providers in which both providers practiced eligible specialties, shared the same HRR, and each billed Medicare for at least 100 office visits. 94.2% of physician pairs were mutually reachable in the patient-sharing networks, and 72.0% were connected on shortest paths of length 4 or less.

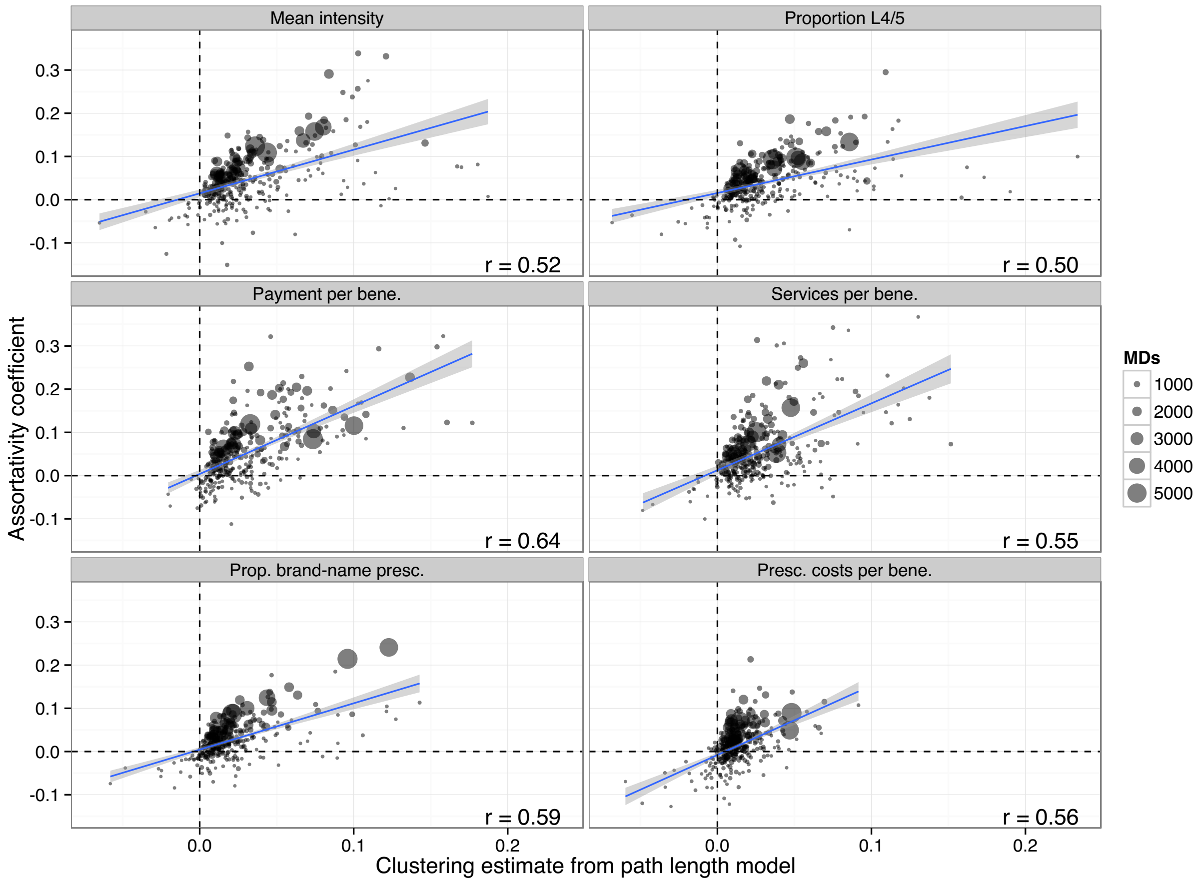

Fig. S5. Comparison of clustering estimates from path length model and assortativity approach. Assortativity on intensity furnishes a nonparametric measure of clustering that accounts for the structure and degree distribution of HRR networks. Estimates of the effect of path length on dyadic difference, by contrast, estimate clustering differently, accounting for physician fixed effects as well as three measures of distance (connected, geodesic=2, and geodesic=3+) rather than two (connected vs. unconnected). For 6 measures of specialty-adjusted intensity and 306 HRRs, 94.0% of clustering estimates from the path length models are positive, while 76.1% of assortativity coefficients are positive. Generally, assortativity estimates are relatively greater for larger HRRs, while estimates from the path length models are greater for smaller HRRs. The two approaches are highly correlated for all measures of billing intensity (Pearson correlation coefficients depicted in figure all significant at $p < 0.0001$).

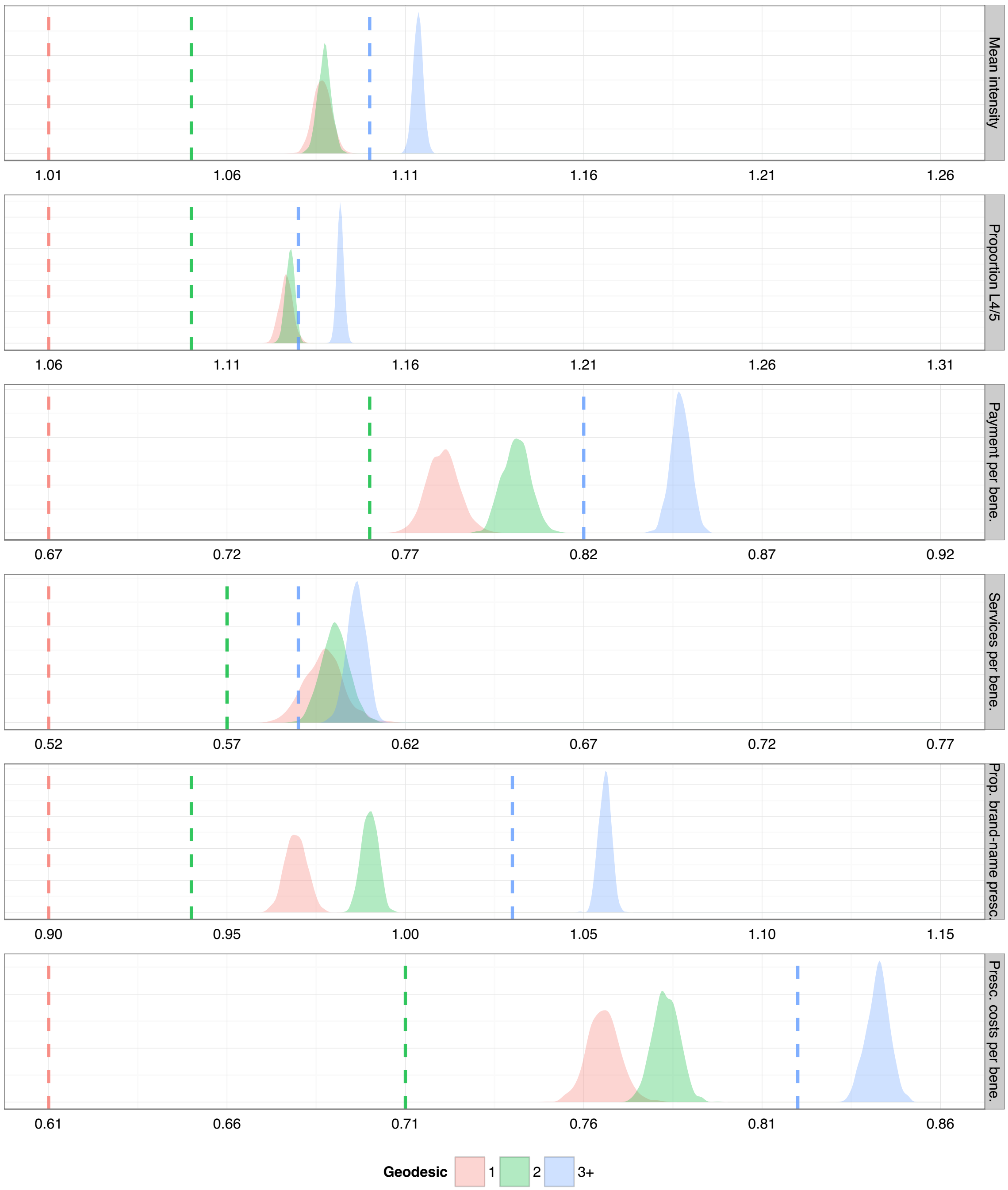

Fig. S6. Pairwise differences in specialty-adjusted intensity measures, by path length. Dashed vertical lines denote the mean pairwise difference in the specialty-adjusted intensity measures, by length of geodesic separating the providers. Smaller values indicate greater similarity in intensity between providers. The distributions represent mean pairwise differences from 1000 simulations, retaining network structure and randomly permuting provider attributes within HRR. Dashed lines further removed from their corresponding null distributions represent results less likely to have arisen due to chance alone.

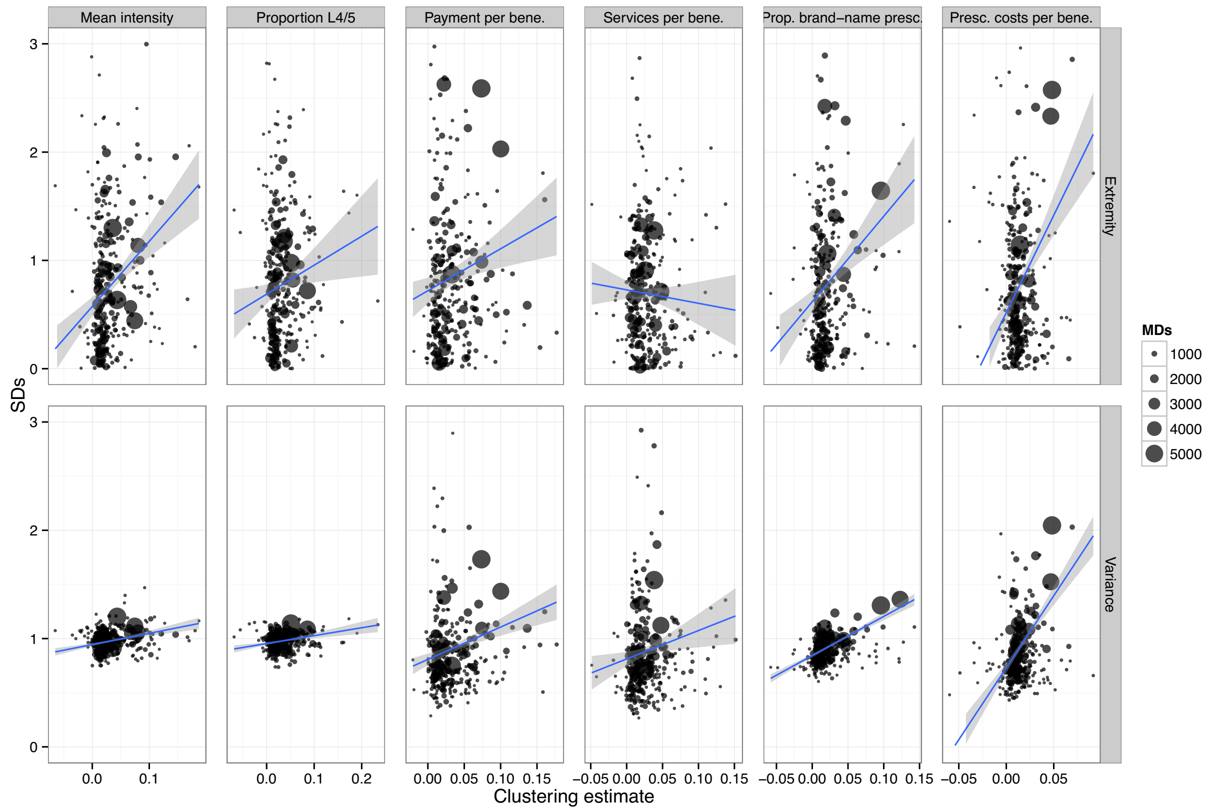

Fig. S7 Greater clustering on intensity is associated with more extreme and varied HRRs. In the top row, the absolute value of the HRR-level z-score for each measure of physician intensity (representing the mean extremity of an HRR's physicians, high or low) is plotted against the clustering estimate for each HRR. In the bottom row, the variance of an HRR's physicians (in SDs) for each measure is plotted against clustering estimates. Lines of best fit represent linear models weighted by the number of physicians in each HRR. For 5 of 6 measures (mean office visit intensity, proportion of high-intensity visits, mean payment per beneficiary, proportion of brand-name prescriptions, and prescription costs per beneficiary), networks exhibiting greater clustering on intensity are more extreme (the HRR-level mean of physician intensities is further from the nationwide average). For all measures, more highly clustered HRRs exhibit greater variance (the physicians within the HRR exhibit greater variability in intensity, $p < 0.05$).

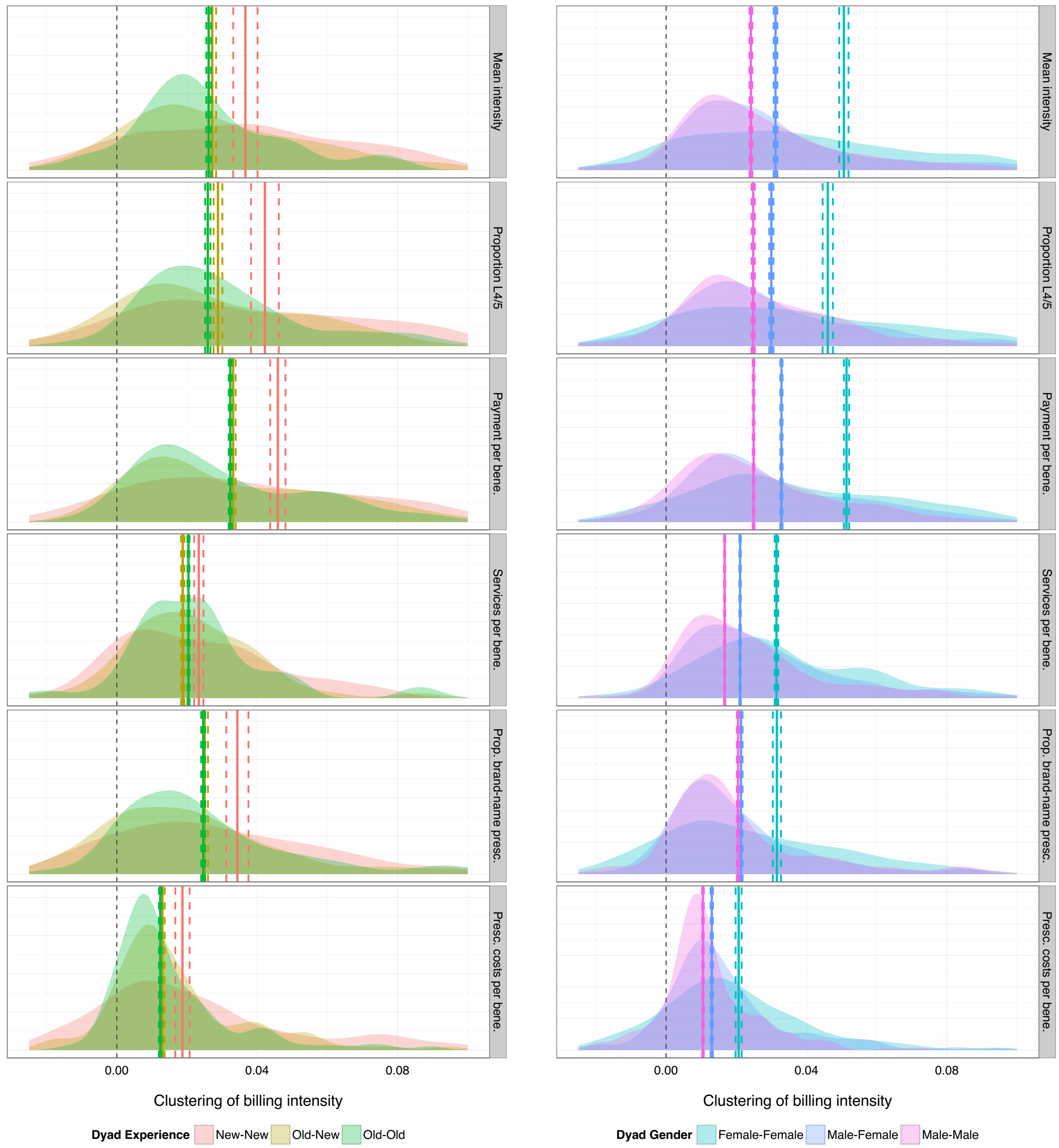

Fig. S8. Estimates of clustering on six measures of specialty-adjusted billing intensity, stratified by dyad experience and gender concordance. Density plots depict the distribution of clustering estimates for each of the United States' 306 HRRs. Solid vertical lines denote pooled nationwide estimates (from fixed-effects meta-analysis of HRR-level results, i.e., larger HRRs weighted more heavily). Corresponding dashed lines indicate 95% confidence intervals. All intensity measures show positive evidence of clustering: controlling for the unobserved attributes of individual physicians, directly connected providers are more alike than pairs of providers separated by a shared colleague, who in turn are more alike than pairs of providers on geodesics of 3 or more. For all measures, significantly greater clustering is manifest among pairs of "new" providers (in which both members are estimated to have been in practice for less than 7 years) than for pairs of "old" providers (both members in practice for 7 or more years) or pairs of mixed experience (left panel). Likewise, greater clustering is observed among pairs of female providers than among mixed-gender or male-male dyads (right panel).

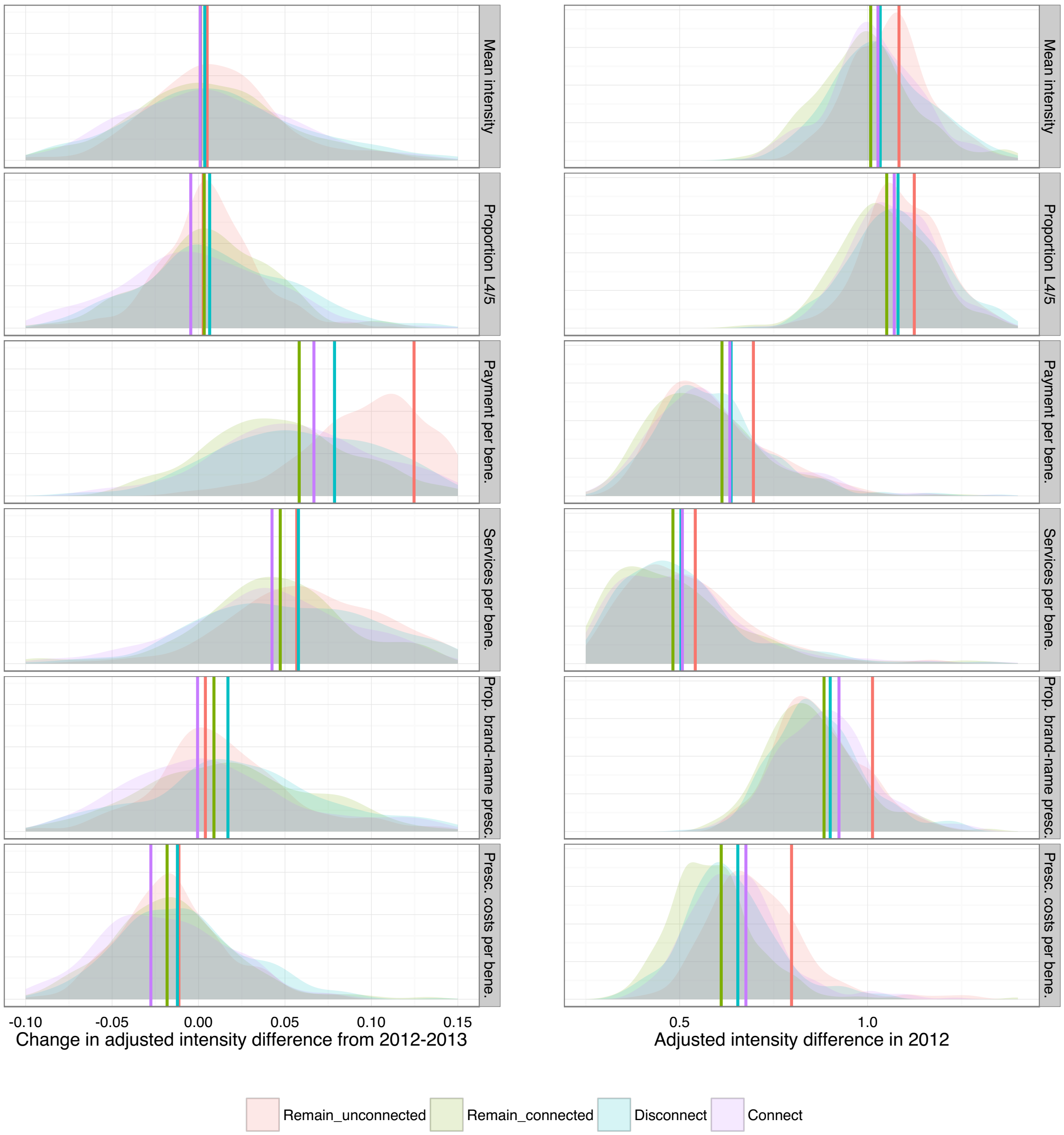

Fig. S9. Nonparametric tests for 'influence' and 'selection.' To test for social influence among physicians as a mechanism contributing to observed clustering on intensity, we evaluate each dyad's *change* in Y_{ij} from 2012-2013, ΔY_{ij}, as a function of the dyad's change in connection status Δc_{ij} (i.e., remain unconnected, remain connected, disconnect, or connect). Density plots depict HRR-level mean ΔY_{ij} for each value of Δc_{ij}; vertical lines denote weighted means of HRR-level results (weighted by number of dyads of each type in each HRR). Consistent with influence, dyads that form patient-sharing connections in 2013 also become more similar in specialty-adjusted intensity than those who remain unconnected (left panel). To test for selection as a mechanism of clustering, we evaluate each dyad's adjusted intensity differences Y_{ij} in 2012, as a function of its change in connection status Δc_{ij}, as above. Consistent with selection, we find for all measures that dyads who were more similar at baseline in 2012 were more likely to form connections in 2013 (right panel). Likewise, dyads that severed connections in 2013 were more dissimilar in 2012 than those who maintained connections.

	Geodesic=1	Geodesic =2	Geodesic =3+
Number of dyads	2,225,560	15,134,239	95,187,434
Proportion of dyads	0.02	0.13	0.85
Difference in mean intensity	1.01 (30.62)	1.05 (20.16)	1.10 (10.31)
Difference in proportion of L4/5 visits	1.06 (36.63)	1.10 (21.80)	1.13 (13.00)
Difference in payment per beneficiary	0.67 (24.79)	0.76 (10.73)	0.82 (10.26)
Difference in services per beneficiary	0.52 (14.10)	0.57 (7.87)	0.59 (6.09)
Difference in proportion of brand-name prescriptions	0.90 (23.11)	0.94 (22.04)	1.03 (15.04)
Difference in prescription costs per beneficiary	0.61 (34.05)	0.71 (19.87)	0.82 (6.97)

Table S1. Pairwise differences in specialty-adjusted intensity measures, by geodesic. Values denote the mean pairwise difference in z-scores for each of the specialty-adjusted intensity measures, by geodesic (shortest path length) separating the providers: smaller values indicate greater similarity in intensity between providers. Values in parentheses denote the number of standard deviations below the mean of the corresponding distribution of permuted values (see Fig. S6): larger values indicate that dyads are more similar than would be expected based on chance and network structure alone.

	Proportion of patient-sharing instances	Proportion of shared patients	Number of patient-sharing instances	Number of shared patients
Difference in mean intensity	-0.06	-0.06	-0.02	-0.03
Difference in proportion of L4/5 visits	-0.05	-0.06	-0.02	-0.04
Difference in payment per beneficiary	-0.05	-0.07	0.05	-0.04
Difference in services per beneficiary	-0.04	-0.07	0.03	-0.06
Difference in proportion of brand-name prescriptions	-0.04	-0.04	-0.03	-0.03
Difference in prescription costs per beneficiary	0.00	-0.01	0.00	-0.06

Table S2. Spearman correlations between measures of relationship strength and pairwise differences in specialty-adjusted intensity measures. Among pairs of providers sharing 11 or more patients, ties constituted by a greater proportion of each provider's shared patients exhibit greater similarity in the billing behavior of the two providers. All correlations are significant at $p < 0.0001$.

	Effect of disconnection on change in specialty-adjusted difference	Effect of specialty-adjusted difference in 2012 on probability of severing connection in 2013
Difference in mean intensity	-0.001 (-0.003, 0.000)	0.013 (0.012, 0.014)
Difference in proportion of L4/5 visits	-0.001 (-0.002, 0.001)	0.010 (0.010, 0.011)
Difference in payment per beneficiary	0.003 (0.002, 0.004)	0.030 (0.028, 0.031)
Difference in services per beneficiary	0.001 (0.000, 0.002)	0.044 (0.042, 0.045)
Difference in proportion of brand-name prescriptions	-0.001 (-0.002, 0.001)	0.013 (0.012, 0.014)
Difference in prescription costs per beneficiary	0.001 (0.000, 0.002)	0.014 (0.012, 0.015)

Table S3. Additional model results for detection of influence and selection. Table entries denote coefficient estimates from the "disconnection" models, using the method of Gaure (27, 28). Values in parentheses denote 95% confidence intervals. For specialty-adjusted payment per beneficiary, dyads that sever patient-sharing relationships between 2012 and 2013 also diverge in intensity over that period (left column). This is potentially consistent with influence, whereby social-professional relationships induce similarity in behavior, which diverges when the relationship is terminated. For all six measures, connected dyads with greater differences in specialty-adjusted intensity in 2012 are more likely to sever their relationships in 2013 (right column). This is consistent with selection, whereby providers preferentially maintain relationships with more similar providers.